

July 8, 2024

1 Conclusiones

Hemos terminado el curso y vamos a resumir sobre lo que es de impacto. - Acabamos de terminar un curso de estadística descriptiva, donde la estadística tiene 2 visiones fundamentales: - Resumir información con números concretos (desviación estándar, promedio, mediana...) y también resumir información con visualizaciones. - Estadística descriptiva como herramienta de análisis nos permite trabar en 2 bloques fundamentales de la ciencia de datos: - **análisis exploratorio de la información** - **procesamiento de la información** antes de tener un modelo de Machine Learning.

1.0.1 ¿Por qué eso es tan importante?

Por que lo que nos encontramos comúnmente en casos reales, es tenemos un data set crudo donde hay datos que pueden tener distintos tipos; pueden ser numérico, categórico. Tenemos que hacer transformaciones sobre los números y sobre las categorías, para que todo tenga un formato estandarizado y homogéneo que sea de fácil entendimiento para un modelo de Machine Learning. Aquí el concepto de correlación es fundamental, entonces usamos las correlaciones para visualizar que variables tienen una alta relación unas con otras y de ahí evidenciar que probablemente hay variables que pueden resultar redundantes para dárselas como input a un modelo de Machine Learning (predictivo o de clasificación).

Entonces todo eso tiene una razón de ser, la idea del curso es precisamente que nosotros adquiriéramos los elementos de estadística descriptiva mínimos para que sepamos todo lo que hay que hacer en procesamiento, que involucra estadística cuando preparamos el data set desde lo crudo hasta la predicción y el uso en un modelo.

Como conclusion final, tenemos entonces la herramienta de análisis de componentes principales que es una herramienta muy poderosa, pero no es la única. La finalidad es dar el impulso para animarnos a explorar e indagar mas sobre que otros métodos existen.